Influence of Table of Specification on the Construction of Ordinary Level Physics Examination in Cameroon

Awandia Joseph Tazitabong Ph.D.[1], **Dikande Alain Moise**[2], **Ndifon Isaiah Ngek** Ph.D.[3]

[1]Department of Curriculum Studies and Teaching, Faculty of Education,
University of Buea, Buea, Cameroon
Assessor, Ordinary Level Physics, GCE Board
[2,3]Department of Physics, University of Buea, Buea, Cameroon

ABSTRACT
A table of specification is fundamental in a test construction. The use of table of specifications when construction a teacher-made achievement test and standardized test is very essential, because it will make the test valid and reliable. Unfortunately, because of lack of inadequate training on its use, it is usually not used by many teachers when constructing a test. The results from these types of assessments are likely not to be valid and reliable. In this situation, some topics that the teacher spent little time in teaching may carry more weighting leading to students' poor performance in the subject (Physics). Most teachers and administrators are still relatively blank as far as skills in test construction and interpretation are concerned. Classroom test provides teachers with essential information that they can use to make decisions about instructions, students learning and student grades. This paper is centred on the following; meaning of weighting, table of specification, the purpose of the table of specification, the benefits of a table of specification in test construction, what should be taken into account when building a TOS, a practical example of TOS, Bloom's taxonomy of educational objectives and item analysis. The importance of table of specifications and the inherent dangers of not using it are highlighted and recommendations to ameliorate the situation are proffered.

KEYWORDS: *Table of specification, test construction, bloom taxonomy, item analysis and examination*

INTRODUCTION
A classroom test plays a central role in the assessment and evaluation of learners. A test provides relevant quantitative information that usually guide critical and crucial decisions about individuals or groups in an institution. The validity of the information depends on the care that was taken in the planning and construction of the test. Since a good test measure what it is meant to measure systematically, it means that there are some systematic steps, principles or procedures involved in test construction. The use of the table of specification in constructing a test ensures that it has high content validity. A Classroom test must be aligned to the content taught in order for any judgment about the student understanding and learning to be meaningful (Alade, & Igbinosa, 2014).

Every classroom assessment measure must be appropriately reliable and valid, be it the classic classroom achievement test, attitudinal measure, or performance assessment. A measure must first be reliable before it can be valid. Classical test reliability and validity must relate to consistent (reliable) and accurate (valid) measurement (Helenrose & Nicole, 2013). Reliability as an indicator of consistency is an indicator of how stable a test score or data is across time. A measure should produce similar or the same results consistently if it measures the same thing. A measure can be reliable without being valid. A measure cannot be valid without being reliable. Some major factors that are a threat to reliability include group homogeneity; when a test is given to a very similar homogenous group, the resulting score are closely clustered making the reliability coefficient, to be low. The more heterogeneous the examined group, the higher the correlation; the time limits; the rate at

which an examinee work will systematically influence performance, as some will finish the test and some will not. Length of the test: if a test is too short, then the reliability coefficient will be low and therefore resulting to scoring errors. All these are threats to reliability of test items constructed which the teachers must take into consideration. Evidence based on test content underscores the degree to which a test measures what it is designed to measure (Wolnring & Wilkstron, 2010).

A content valid test should have at least moderate to high levels of internal consistency. This suggest that the items measure a common element; primarily rest upon logical argument and expert judgment, and frequently empirical research. The degree of content validity is largely a function of the content to which test items are true representative samples of the content and skills to be learned (Onunkwo, 2002; Wolnring & Wikstron, 2010). Standardized test scores are frequently different among students' GPA and scores on a standardized test, sometimes very large differences from the literature. We know standardized tests are valid. The question needs to be asked if GPAs are valid measures of students' achievement. This is because, GPAs are based on teacher made tests. If teacher made tests are not valid, how can a student GPA be valid? The use of the table of specification can ensure that a teacher made test is valid. For validity to be achieved, the test designer must first of all start by considering the weighting of the various topics.

Weighting refers to the assignment of numerical values (marks, scores or percentages) to test items or questions. In terms of syllabus, it may also refer to the assignment of percentages to various test items or sections of the syllabus or to each paper or section of an examination or test. Weighting is also done in terms of objectives, content and different forms of questions. It is usually done with respect to the cognitive levels and levels of difficulty or number of skills involved. To make a test valid, it is necessary to analyze the objectives of the subject and decide which objectives are to be tested and in what proportion. Marks should be allotted to each objective to be tested according to its importance. In physics testing at the ordinary level in Cameroon, the four cognitive abilities tested are; knowledge of the subject matter, comprehension, application and analysis. The weighting to all these four may be decided in percentages, for example for a test of 50 marks for ordinary level the following weightings may be used as shown on table 1.

Table 1: the cognitive ability, percentage of marks and marks allotted in the physics testing at the GCE ordinary level in Cameroon

Cognitive ability	Percentage of Marks	Marks allotted
Knowledge	30%	15
Comprehension	40%	20
Application	20%	10
Analysis	10%	05
Total	100%	50

It is also necessary to analyze the syllabus and allot weighting to different areas of content. This is again done to ensure the validity of the test. A hypothetical example is given below for a physics test showing weighting to content units for a class test, this is illustrated on table 2.

Table 2: the content area, percentage of marks and marks allotted.

Content Area	Percentage of Marks	Marks allotted
Heat	30%	30
Electricity	40%	40
Waves	30%	30
Total	100%	100

After analyzing the objectives and the content, the next step is to decide how they will be tested. A particular objective and content can be tested more appropriately by a particular form of question. So, different forms of questions are to be included in the test for testing different objectives and contents. For this reason, the number of different types of questions to be included in the test and the marks carried by each of them are decided. This takes care of the reliability of the test. As an illustration, hypothetical weighting to different forms of questions in paper one and two in a typical physics test for form five is given on table 3

Table 3: forms of questions, no. of questions, marks allotted and % of marks for a typical form five test.

Forms of Questions	No. of Questions	Marks allotted	% of marks
Short Type	6	25	25%
Essay Type	3	35	35%
MCQ	50	40	40%
Total		100	100%

When all the above activities and assessments have been achieved, the next logical step is to design the table of specification.

The Table of Specification is a plan prepared by the examiner or test item developer as the basis for test construction. It is a tool that teachers and examiners may use in test construction to mitigate the problem of mismatched assessment. The table of specification (TOS), is also referred to as test blue print (TBP). It is a table that helps teachers and examiners to align objectives, instructions and assessment (Zuelk, Wilson and Yunker, 2004). Gregory (2006) says the TOS is an activity which enumerates the information and cognitive tasks on which examinees are to be assessed. It is a chart that professional developers of achievements and ability tests often use in test item writing. Onunkwo, (2002); Wolnring, & Wikstron, (2010) stated that the TOS or test blue print is a device that enables the teacher (examiner) to arrive at a representative sample of the instructional objectives and the subject matter treated in the class and what is covered in the assessment. The TOS is a guide to assist the teacher or examiner in the evaluation process. It is developed from the content of a subject or curriculum that is broadly defined to include both subject matter content and instructional objectives.

A Table of specification for practical classroom application is intended to help classroom teachers develop summative assessments that are well aligned to the subject matter studied and the cognitive process used during instruction. However, for this strategy to be helpful in your teaching practice, you need to make it your own and practical assessment. Gronlund and Linn (2000) assert that a table of specification may be referred to as content of a course or curriculum that can be broadly defined to include both subject matter content and instructional objectives. This simply means the performance of students is expected to demonstrate both of these aspects (Grunlund, 2000; Onunkwo, 2002; Wolnring, & Wikstron, 2010). Akem (2006) views the table of specification as a guide to assist a teacher or examiner in the evaluation system.

A table of specification shows the total number of items to be allocated to each instructional objective, it also suggests what might be covered under each item, take decision on what type of items to be used. In fact, the blue – print stage is the last and crucial stage in an evaluation plan since it enables the teacher to combine properly the objectives and the content areas, bearing in mind the importance and the weight attached to each area. Akem and Agbe (2003) revealed that a table of specification is an outline relating behaviour to topics. By it, teacher can determine what topics are being stressed and also assist in the preparation of test that reflect what students have learned and also the amount of time

spent on each unit. Okpala, Onocha and Oyedeji (2003) noted that a table of specification enables the test developers to complete the cells in the table and decide the percentage of the total number of items that will go to each of the cell. Ughamadu (2000) stated that a table of specification or test blue print is a device that enables the teacher to arrive at a representative sample of the instructional objectives and the subject matter treated in the class.

Importance of Table of Specifications (TOS)
The most important purpose of the TOS is to achieve a balance in a test; by identifying achievement domains being measured, and to ensure that a fair and representative sample of questions appear on the test. (Remember it is impossible to ask questions on every aspect and objective of a syllabus in one examination). The second important purpose is to ensure that our test focuses on the most important areas of the syllabus or curriculum, and weights different areas based on their importance and time spent in teaching. The third purpose is that TOS provides proof that our test has content validity, which significantly covers the syllabus. It ensures that the test is within the prescribed level of the learners or those to be assessed as indicated in the syllabus (Helenrose & Nicole, 2013).

The benefits of the table of specification in test item construction include the following: The TOS ensures that an assessment has content validity. That means it tests what it is supposed to test; there is a match between what is taught and what is tested. It ensures that the same emphasis on content is mirrored on assessment. This means topics which are more important would have more items. It ensures alignment of test items with the objectives of the syllabus (e.g unimportant topics may just test knowledge, while important topics would test interpretation, application and synthesis).

The purpose of a table of specifications is to identify the achievement domains being measured and to ensure that a fair and representative sample of questions appears on the test, thereby improving the validity of teacher's evaluation based on a given assessment. The importance of table of specifications as a guide to test construction cannot be over emphasized as opined by Denga (2003). Thus:

➤ It defines as clearly as possible the scope and emphasis of the test, to relate the objective to the content and to construct a balanced test.

➤ Through the use of table of specifications, teachers are able to determine what topic is being stressed and also assist in the preparation of tests that reflect what students have learnt and also limit the amount of time spent on each unit.

➢ It constrains the tester and ensures that only those objectives involved in the instructional process are assessed. There is a balance in testing the materials taught because each objective receives proportional emphasis in relation to the amount of time given it and the value placed on it.

➢ It helps the teacher in organizing teaching and learning, assessment and evaluation as well as all the resources he plans to achieve during teaching and learning.

➢ It assists immensely in the preparation of test items, production of the valid and well robust test, in the classification of objectives to both teacher and students, and in assisting the teacher to select the most appropriate teaching strategy

Defects of not using table of specifications in test construction

According to Ehiagwina (2019), The test so prepared without a table of specifications will lack content validity. The scores obtained from such a test are not a true representative of the pupils/students actual subject standing, since all the topics are not covered. The pupils/students might be denied the areas they will have performed excellently and given the area he/she could not perform well. There will be errors in placement and interpretation of student's actual physics performance. The test items that lack a table of specification might not tie with the test taker's cognitive level. It might be below or above the test taker's cognitive ability.

Also, in order to construct a table of specification, or test blue print, which will adequately guide in developing a test that truly represents its content and objectives, Nenty and Imo (2004), Joshua (2005), pointed out the following steps in the preparation of a table of specification:

1. Decide on the total number of items that will constitute the test

2. Decide on the percentage of items to be prepared on each content topic or unit

3. Decide on the percentage of items to be prepared on each level of the instructional objectives (cognitive domain)

4. Determine the actual number of items to be prepared on each content topic/ unit (i.e, the row totals) using the number and percentages specified in steps (1) and (2).

5. Determine the actual number of items to be prepared on each level of the instructional objective (i.e, the column totals) using the number and percentages specifies in steps (1) and (3)

6. Determine the actual number of items to be prepared on each content topic/ unit for the different cognitive levels (i.e, filling the cells in the body of the table) using the specified percentages and the row and column totals

7. Make the necessary minor adjustment if any (i.e, rounding up of decimal points), but ensure that the row and column totals are maintained.

The procedure for developing a good test (or factors to consider in the construction of a good test) according to Nenty and Imo (2004) and Joshua (2005) consist of the following systematic steps:

a. Specify the purpose (goals or objectives) of the test

b. Develop a test blue - print or table of specification

c. Develop test items

d. Select the items

e. Prepare test instructions

f. Assemble the test

g. Due preliminary administration of the test

h. Determine the reliability of the final test

i. Determine the validity of the final test

j. Print the final copy after editing and proof reading have been done.

Some of the reasons or purposes for testing according to Nenty and Imo (2004); Helenrose and Nicole (2013), are to;

1. Evaluate the teacher's instructional method

2. Ascertain the effectiveness, validity and level of coverage of a curriculum

3. Motivate students

4. Judge the pupils' mastery of certain essential skills and knowledge

5. Diagnose students' difficulties

6. Rank students in terms of their achievement of particular instructional objectives

7. Measure growth overtime. The teacher usually starts the term by specifying the course or subject instructional objectives. That is, those specific things the students/ pupils will be able to accomplish at the end of instructional period.

There are three main steps involved is preparing instructional objectives. These are;

a. Identifying the general instructional objectives

b. Stating the general instructional objectives

c. Defining the general instructional objectives.

The purpose is to coordinate the assessment questions with the time spent on any particular content area, the objectives of the unit being taught, and the level of critical thinking required by the objectives or state standards. Tables of Specifications are created as part of the preparation for the unit, not as an afterthought the night before the test. Knowing what is contained in the assessment and that the content matches the standards and benchmarks in level of critical thinking will guide learning experiences presented to students. Students appreciate knowing what is being assessed and what level of mastery is required.

According to Moore (2001), a good instructional objective has four components namely;

1. Performance statement - which indicates the specific behaviour the leaner will be able to show or exhibit. It must be stated in terms of what students are expected to do. That is, observable students' performance. So proper verbs must be used.

2. Product – What students will produce by action. It is this product that will be evaluated to determine whether the objectives have been mastered. E.g., a written statement, sum, listed names, etc.

3. Condition statement-Which indicates the conditions under which the performance statement (expected behaviour) is expected to occur. This is usually when the teaching exercise has been completed. For example, will they be allowed to use open book, will material be provided, etc.

4. Criterion statement -Which indicates the level or standard of performance (behaviour) that will be acceptable. What is the level of acceptable student performance? Here, the level of behaviour that will be accepted as satisfactory must be stated.

The following is an example of a well-stated instructional objective in physics at the ordinary level;

At the end of this instructional exercise the students should be able to state correctly the three Newton's laws of motion.

Performance statement-State the three Newton's laws of motion

Product statement – The three laws written or stated.

Condition statement-After attending the instructional session

Criterion statement - Each student will state the laws correctly

Instructional objectives define the course content (topics) to be selected, the purpose of the test to be given and the content of the test items to be developed. A good teacher must be well versed in the development and appropriate stating of instructional objectives.

Table 4 shows the general format of a table of specification. It is a two-dimensional table that relate levels of instructional objectives to the subject or topic content. That is, it guides a test constructor in the selection of items. It is a systematic procedure of ensuring that the instrument (the test) adequately covers all the behavioural domains to be measured in relation to the programme content. The level of instructional objectives in the cognitive domain are arranged at the top (in columns) and the typical or unit in the course content are arranged vertically to the left (in rows). It contains the number of items to be set from each section of the subject content per cognitive level.

Table 4: the general format of a table of specification

Content	Knowledge percentage	comprehension percentage	Application percentage	Analysis	Total
Topic 1					
Topic 2					
Topic 3					
Topic 4					
Topic 5, etc.					
Total					

In each cell, the number and / percentage of item to be constructed are indicated. This depends on the relative emphasis on topics and behaviours as might be indicated by the instructional objective. For example, if a teacher wants to develop an end of term test in physics, he may have to consider the following; course objectives, topics covered in class, amount of time spent on those topics, and emphasis and space provided in the text.

A sample of a table of specification for a 50- item objective test for physics 0580 GCE ordinary level is given on table 5. A sample TOS for a 50-item objective test for 0580 physics, with knowledge 30%, Comprehension 40%, application 20% and analysis 10 % is shown on table 5;

Table 5: the table of specification (Test blue print) for an ideal physics – 0580 paper one

S/N	Content	Knowledge (30%)	Comprehension (40%)	Application (20%)	Analysis (10%)	Total (100%)
1	Forces (6) =12%					
2	Motion (4) = 8%					
3	Energy (3) = 6%					
4	Heat (6) = 12%					
5	Properties of Matter (4) = 8%					
6	**Electricity (9) = 18%**					
	-Electrostatics (3) = 6%					
	-Current Electricity (6) = 12%					
7	Electromagnetism (4) = 8%					
8	**Modern Physics (6) = 12%**					
	-Electronics (1) = 2%					
	-Nuclear Physics (5) = 10%					
9	**Waves (8) = 16%**					
	-Optics (3) = 6%					
	-Waves Properties (5) = 10%					
	Total (50)	15	20	10	5	50

Table 5 shows that of the 50 items of the test; 15 (30%) will be based on knowledge or memory, 20 (40%) will be on comprehension, 10 (20%) will be on application and 5 (10%) will be based on analysis. It also shows that 6 (12%) will be based on forces, 4 (8%) on motion, 3(6%) on energy, 6 (12%) on heat, 4 (8%) on properties of matter, 9 (18%) on electricity, 6 (12%) on modern physics and 8 (16%) on waves.

The percentages on the rows and columns are usually used to fill the table as follows;

To determine the total number of questions under each behavioural objective, we use the percentage of the total number of questions. E.g.;

For knowledge
In the table above 30% of 50 items equals; $\frac{30}{100} x \frac{50}{1} = 15 \ items$

For comprehension
40% of 50 items equals; $\frac{40}{100} \times \frac{50}{1} = 20 \ items$

For application
20% of 50 items equals; $\frac{20}{100} \times \frac{50}{1} = 10 \ items$

For Analysis
And 10% of 50 items equals; $\frac{10}{100} \times \frac{50}{1} = 5 \ items$

All these are shown on the row for total under each heading. In the same way content can be computed as shown below:

1. Forces = 12%
12% of 50 items equals: $\frac{12}{100} \times \frac{50}{1} = 6$ items for forces

2. Motion = 8%
8% of 50 items equals: $\frac{8}{100} \times \frac{50}{1} = 4 \ items$ for motion

3. Energy = 6%
6% of 50 items equals: $\frac{6}{100} \times \frac{50}{1} = 3 \ items$ for energy

4. Heat = 12%

12% of 50 items equals: $\frac{12}{100} \times \frac{50}{1} = 6 \, items$ for heat

5. Property of Matter = 8%

8% of 50 items equals: $\frac{8}{100} \times \frac{50}{1} = 4 \, items$ for properties of matter

6. Electricity = 18%

18% of 50 items equals: $\frac{18}{100} \times \frac{50}{1} = 9 \, items$ for electricity

7. Electromagnetism = 8%

8% of 50 items equals: $\frac{8}{100} \times \frac{50}{1} = 4 \, items$ for electromagnetism

8. Modern Physics = 12%

12% of 50 items equals: $\frac{12}{100} \times \frac{50}{1} = 6 \, items$ for moderm physics

9. Waves = 16%

16% of 50 items equals: $\frac{16}{100} \times \frac{50}{1} = 8 \, items$ for waves

This completes the column under total. Both sides should sum up to 50 by adding down wards and across the total column (for content areas and behavioural objectives respectively)

To complete the inside (cells) we can either use the column totals or the row totals both of which should give the same result if computed correctly using the column totals.

Beginning with the first column;

12% of 15 = $\frac{12}{100} \times \frac{15}{1} = 1.8$, 8% of 15, = $\frac{8}{100} \times \frac{15}{1} = 1.2$, 6% of 15 = $\frac{6}{100} \times \frac{15}{1} = 0.9$

12% of 15 = $\frac{12}{100} \times \frac{15}{1} = 1.8$, 8% of 15, = $\frac{8}{100} \times \frac{15}{1} = $, 1.2, 18% of 15 = $\frac{18}{100} \times \frac{15}{1} = 2.7$

8% of 15, = $\frac{8}{100} \times \frac{15}{1} = $, 1.2, 12% of 15 = $\frac{12}{100} \times \frac{15}{1} = $, 1.8 and lastly, 16% of 15 = $\frac{16}{100} \times \frac{15}{1} = 2.4$

These are the entries in the cells under knowledge. Summarily, the entries under comprehension, application and analysis can be done the same. At the end we will come up with a table like that shown on table 6.

Table 6: a raw table of specification (test blue print) for an ideal physics – 0580 ordinary level

S/N	Content	Knowledge (30%)	Comprehension (40%)	Application (20%)	Analysis (10%)	Total (100%)
1	Forces (6) = 12%	1.8	2.4	1.2	0.6	6
2	Motion (4) = 8%	1.2	1.6	0.8	0.4	4
3	Energy (3) = 6%	0.9	1.2	0.6	0.3	3
4	Heat (6) = 12%	1.8	2.4	1.2	0.6	6
5	Property of Matter (4) = 8%	1.2	1.6	0.8	0.4	4
6	**Electricity (9) = 18%**	**2.7**	**3.6**	**1.8**	**0.9**	**9**
	-Electrostatics (3) = 6%	0.9	1.2	0.6	0.3	3
	-Current Electricity (6) = 12%	1.8	2.4	1.2	0.6	6
7	Electromagnetism (4) = 8%	1.2	1.6	0.8	0.4	4
8	**Modern Physics (6) = 12%**	1.8	2.4	1.2	0.6	**6**
	-Electronics (1) = 2%	0.3	0.4	0.2	0.1	1
	-Nuclear Physics (5) = 10%	1.5	2.0	1.0	0.5	5
9	**Waves (8) = 16%**	**2.4**	**3.2**	**1.6**	**0,8**	**8**
	-Optics (3) = 6%	0.9	1.2	0.6	0.3	3
	-Waves Properties (5) = 10%	1.5	2.0	1.0	0.5	5
	Total (50)					**50**

The next step is to convert the above table into a practical one by rounding up the fractions taking into consideration the level of the students or candidates.

Table 7: table of specification (Test blue print) for physics – 0580 completely filled only with whole numbers.

S/N	Content	Knowledge (30%)	Comprehension (40%)	Application (20%)	Analysis (10%)	Total (100%)
1	Forces (6) = 12%	2	2	1	1	6
2	Motion (4) = 8%	1	2	1	-	4
3	Energy (3) = 6%	1	1	1	-	3
4	Heat (6) = 12%	2	2	1	1	6
5	Property of Matter (4) = 8%	1	2	1	-	4
6	Electricity (9) = 18%	3	4	1	1	**9**
7	Electromagnetism (4) = 8%	1	2	1	-	4
8	Modern Physics (6) = 12%	2	2	1	1	**6**
9	Waves (8) = 16%	2	3	2	1	**8**
	Total (50)	**15**	**20**	**10**	**05**	**50**

When a table of specification that shows how many items are to be constructed on each cognitive level for each topic on the subject content being determined has been developed, the next logical activity is to develop or construct the items based on those specifications. Item's development means translating the subject content into test items (questions or statements) that will stimulate the test takes and elicit the type of behaviour specified in the subject instructional objectives. Test measures behaviour or attribute indirectly. For example, items development implies writing statements that call for specific behaviours of the test takers and the test takers responses to the item and these behaviours will indicate the amount or level of the trait being measured or the amount of the content that has been mastered.

According to Nenty and Imo (2004) classroom test item can be categorised into two, namely;
1. Objective items which are highly structured and require the examinees to supply a word or two or to select the correct answer from a number of alternatives
2. The essay items which allow the examinees to supply, organize and present the answer in essay form.

The use of each type is based on;
a. The learning outcomes to be measured
b. The advantages and limitations of each type
c. The level of maturity of the testees
d. The skills of the test developer.

The number of items to be developed is based on;
1. How many items are needed to ensure satisfactory reliability and content coverage or content validity.
2. How can these important but internal test characteristics be skilfully balanced against the many external constraints on the length of the test?
3. How many items should be written initially to ensure that a sufficient number would survive item review and analysis after try-out? So, the test developer starts by writing more items than the number needed. The first draft is then reviewed and edited by correcting ambiguous wordings, strengthening weak alternatives and eliminating duplicates and otherwise unsuitable items.
4. The next step is to select the correct items guided by the table of specifications. For standardized test, items selection is carried out after item analysis so that the selection will take into consideration the levels of difficult and discrimination between the bright and the slow learners.

It has been noticed that, for teacher - made test for use in the classroom or specific school the items are hardly tried out and analysed before use. In the absence of formal items try-out and analysis item selection should depend on the result of critical and thorough review and editing of each of the items by senior colleagues in the subject matter and one with expertise in measurement and evaluation. Item selection should be such that only the

learners who have the specified knowledge, ability or characteristics being measured can respond correctly to them and no other ability or characteristic should influence the learners in their performance.

Assembling the test

According to Joshua (2005); Nenty and Imo, (2004) assembling of the test can be subdivided into two parts, namely for teacher - made test and for standardised test. For teacher made test, the test should be produced in such a way that it can go round to all the testees. The production should be neat and legible and no examinee should be disadvantaged in any way as a result of poor printing or photocopy, wrong spellings, omitted parts of some questions and such other experiences that can affect the testees negatively. Items should be arranged in such a way that there will be rapport between the testees and the test at the beginning.

For standardised test, assembling the test means preparing the final form or forms of the test using the results of item analyses. Items with best discrimination, appropriate difficulty and good distractors are coupled and printed to form a test. There should be balancing and compromising concerning item properties so that all categories of test takers are welcomed or accommodated in the test. Some examination boards like JAMB in Nigeria use about four to five different forms of the same test so as to reduce examination malpractices. But it is doubtful whether the same level of test anxiety, rapport and test difficulty are maintained in the test takers as they face the different forms/types during the testing session (Nenty and Imo, 2004; Joshua, 2005).

Preparation of test instructions

Two sets of directions or instructions are usually required: one for test takers and the other for the test administrator. Directions to the test takers should indicate the nature of the desired responses, and how and where to make the expected responses. The directions should indicate in relatively simple language the purpose of the test, the time limit, the method of recording answers, the way the test is to be scored and whether or not examinees should guess the answers when they are in doubt or do not know (Nenty and Imo, 2004; Joshua,2005; Reynolds, Livingston, and Wilson, 2006).

For the test administrators, the direction should be such that they will be able to explain the rationale for testing procedures including details about arrangement of testing site(s), distribution and collection of test material, timing and how to handle expected problems and questions during the testing session. It is important to note that insufficient or ambiguous instruction(s) or no instruction create(s) confusion and anxiety and can divert examinees concentration, time and energy during the examination. These short comings can compromise objectivity in testing and so it must be taken seriously.

According to Nenty and Imo, (2004); Joshua (2005), test item analysis is the act of 'testing the test items' so as to verify whether each is serving the purpose of testing. The result of item analysis helps in judging the quality of each item, and, thus, helps in improving the item and the skills of the teacher in test construction. The results also provide diagnostic values that could help teachers in planning future learning activities for the learners, and also feedback to students as regards their performance on each item. There are three indices involved in item analysis. These are: Item difficulty; Item discrimination and Option distraction.

Item difficulty is the proportion of test takers who respond correctly to the item. Thus, item difficulty (P-value) is equal to the number of students who score that item right divided by the number of students who attempted the item. P-values vary from zero (0) - for a very difficult item (nobody got it right) to one (1) for a very easy item (everybody got it right). Thus, the higher the difficulty index of the item, the easier the item and verse versa (Joshua, 2005; Reynolds, Livingston, and Wilson, 2006).

Formula of P- value

$$\text{P- value} = \frac{N^o \text{ of students who get item correct}}{N^o \text{ of students who attempt the item}}$$

Table 8: the interpretation of item difficulty index (Nenty and Imo, 2004; Joshua,2005)

Percentage Range	Difficulty Index	Interpretation
71%-100%	0.75 – 1.00	Easy
61%-70%	0.25 – 0.75	Average and needs review
40% - 60%	0.40 – 0.60	Good
30% - 39%	0.30 – 0.39	Fair and needs review
0% - 29%	0.00 – 0.29	Hard needs to be discarded

Example: In a class of 50 students writing a test, 30 students got item 1 correct, the difficulty index of item 1 will be;

P-Value= $\frac{30}{50}$ = 0.60. From table 8, the item was a good one.

For an essay test, item difficulty is found by dividing the mean of all the testees' scores on the item by the maximum score allocated by the examiner to that item.

Formula or P-value for essay test

Item Difficulty (P) = $\frac{\textit{The mean score of all testees on the item}}{\textit{The maximum possible score for the item}}$

Example: If ten students attempt a question whose total score is 10 marks and score the following; 6, 8, 7, 4,3, 5,5,7, 3,8. The P-value will be determined as follows:

Mean Score = $\frac{56}{10}$ = 5.6

P -value = $\frac{5.6}{10}$ = 0.56, this means that the difficulty index is 0.56. From table 8, the item was a good one.

Item discrimination indicates the extent to which an item is able to distinguish (or discriminate) between the more knowledgeable (bright) students and the less knowledgeable (slow) ones.

Formula For calculating item discrimination (D-value)

$$\frac{N^0 of \ bright \ students \ who \ got \ the \ item \ right - N^0 of \ slow \ students \ who \ got \ item \ right}{N^0 of \ students \ in \ each \ group \ (one \ of \ the \ groups)}$$

In calculating item discrimination (D-value), the entire class will be divided equally into bright, average and dull(slow) students' groups.

Steps for estimating item discrimination

Arrange all the scored scripts or papers in order. Assuming there are 50 testees and hence 50 scripts. Arrange them starting with the one with the highest score and the one with the lowest score. Starting from the highest, count the first one -third of the scripts and the last one- third of the scripts from below the pile. $\frac{1}{3}$ of $50 = 16\frac{1}{3}$, so, one should count the first 16 scripts from above and the last 16 scripts below. D-value varies from -1.00 to + 1.00. The higher the index, the better is the item. A negative index indicates that, the item is a bad one which discriminates in the opposite way. That is, more of dull (slow) students than bright ones got that item right. Generally, items with very low, zero, or negative discrimination indices need careful examination and review.

Table 9: *table showing the interpretation of D- value* (Nenty and Imo, **2004;** Joshua,2005)

Discrimination Index	Interpretation
0.30 and above	Good
0.10 – 0.29	Fair
Equal to 0	No discrimination. All students got the item right.
Negative	Poor. The item was flawed or mis keyed.

Option Distraction

A good distractor in a multiple-choice item is one that attracts or distracts more of dull (slow) students than bright students. The distraction power of an incorrect option (distractor) in a multiple choice item is the ability of that option to differentiate between those who do not know (dull or slow ones) and those who know (bright ones). Option distraction indices vary from -1.00 to + 1.00. A high positive index is desirable. A negative index shows that the distractor attracted (or distracted) more of bright students than it did to the dull (slow) ones which is abnormal. Thus, options with very low, zero and negative indices need review.

Formula for Option Distraction

$$\frac{N^0 of \ dull \ (slow) students \ who \ chosed \ the \ option - N^0 of \ bright \ students \ who \ chosed \ the \ option}{N^0 of \ students \ in \ each \ group \ (one \ of \ the \ groups)}$$

Table 10: the Interpretation of option distraction efficiency *(Nenty and Imo, 2004;* **Joshua,2005***)*

DE	Interpretation	Proposed Action
$\leq 39\%$	Non-functional Distractor	Revise or Discard
40% or more	Functional Distractor	Retain

Table 10 shows that, when the DE is less than or equal to 39% the distractor is non- functional but when it is 40% and above it is functional and so it should be retained.

Example 2 shows how to calculate and interpret item difficulty, discrimination index and option distraction efficiency**.** This table shows the answers that were selected by sixty students in one question in a test (Joshua, 2005).

Table 11: how to calculate and interpret item difficulty, discrimination index and option distraction efficiency

Option	Bright	Average	Dull(slow)	Total	p-value	d-value	Distraction index
A	1	7	4	12			
B	0	3	8	11			
C*	18	10	8	36			
D	1	0	0	1			
Total	20	20	20	60			

Calculate
1. The difficulty index.
2. The discrimination index of the test item.
3. The distraction index of each option.

Interpretation the values you have calculated above.

Table 12: how to calculate item difficulty, discrimination index and option distraction efficiency

Option	Bright	Average	Dull (slow)	Total	p-value	d-value	Distraction index
A	1	7	4	12			$\frac{4-1}{20}=0.15$
B	0	3	8	11			$\frac{8-0}{20}=0.4$
C*	18	10	8	36	$\frac{36}{60}=0.6$	$\frac{18-8}{20}=0.5$	Correct Answer
D	1	0	0	1			$\frac{0-1}{20}=-0.05$
Total	20	20	20	60			

Interpretation
Item difficulty is good, because 60% of examinees got it right
Item discrimination is good, because Index is positive and high.
Option A is not too good, because distractor index is low, though positive.
Option B is a good distractor, because index is positive and high.
Option D is a bad distractor, because index is negative, meaning that it distracted bright students instead of dull(slow) ones.

Item analysis is an important phase in the development of a test . In this phase, statistical methods are used to identify test items that are not working well (Notar, Zuelke, Wilson, & Yunker, 2004; Joshua,2005). If an item is too easy, too difficult, failing to show a difference between skilled and unskilled examinees, or even scored incorrectly, item analysis will reveal it. The two most common statistics reported in an item analysis are the item difficulty, which is a measure of the proportion of examinees who responded to an item correctly, and the item discrimination, which is a measure of how well the item discriminates between examinees who are

knowledgeable in the content area and those who are not. An additional analysis that is often reported is the distractor analysis. The distractor analysis provides a measure of how well each of the incorrect options contributes to the quality of a multiple-choice item. Once the item analysis information is available, an item review is usually conducted (Notar, Zuelke, Wilson, & Yunker, 2004); Joshua, 2005).

Once the item analysis data are available, it is useful to hold a meeting of test developers, psychometricians, and subject matter experts. During this meeting the items can be reviewed using the information provided by the item analysis statistics. Decisions can then be made about item changes that are needed or even items that ought to be dropped from the exam. Any item that has been substantially changed should be returned to the bank for pretesting before it is again used for testing. Once these decisions have been made, the exams should be rescored, leaving out any items that were dropped and using the correct key for the items that were found to have been mis-keyed. This corrected scoring will be used to mark the examinees answers (Joshua, 2005; Reynolds, Livingston, and Wilson, 2006).

It must be appreciated that a complete table of specification should cover all the six major categories in the cognitive domain as identified by Benjamin Bloom and his colleagues (1956). For beginners, however, the table of specification may exclude the higher order categories since they are not expected to acquire such skills at that stage of their academic development. Cognitive domain refers to the domain which deals with the "recall or recognition of knowledge and the development of intellectual abilities and skills" (Bloom, 1956).

Benjamin Bloom, et. al. (1956) classified all educational objectives into three, namely: cognitive, affective and *psychomotor* domains. Cognitive domain involves remembering previously learnt matter. Affective domain relates to interest, appreciation, attitude and value. Psychomotor domain deals with motor and manipulative skills. The focus of 0580 -Physics assessment is on the first four categories of the cognitive domains (knowledge, comprehension, application and analysis) as shown on the TOS. As a reminder, these areas of the cognitive domain are reproduced on table 13 with some of their verbs:

Table 13: shows the first four categories of the cognitive domain with their verbs

Category	Description	Keywords (verb)
Knowledge	Recall information	define, label, list, match, name, recall, recognize, reproduce, select, state, quote, recall, write
Comprehension	Understand the meaning, translation, interpolation, and interpretation of instructions and problems. State something in one's own words	comprehend, convert, distinguish, estimate, explain, extend, generalize, give an example, interpret, paraphrase, differentiate, rewrite, summarize, translate, defend, describe, restate, contrast, discuss.
Application	Use a concept in a new situation or unprompted use of an abstraction. Apply what was learned in the classroom into novel situations in the work place	apply, change, compute, construct, demonstrate, discover, manipulate, modify, operate, predict, prepare, produce, relate, show, solve, calculate, illustrate, use, determine, model, perform, present.
Analysis	Separate material or concept into component parts so that its organizational structure may be understood. Distinguish between facts and inferences	analyze, breakdown, compare, contrast, diagram, deconstruct, differentiate, infer, outline, select, separate classify, categorize, subdivide, criticize, simplify, associate, discriminate, identify,

CONCLUSION

Students always complain that teacher-made tests are characterized by over testing, time spent for administration was too short, the test items do not cover the course content. All these show that the test lacked content validity. Constructing fair tests that give accurate information about students learning is an important skill for teachers. The table of specification is often useful to organize the planning process of designing a test which allows the teacher to determine the content of the test. Using TOS to organize a teacher-made test helps to alleviate the content validity problem because it helps the teacher to create a good balance in several areas. (Nenty, 2007; Reynolds, Livingston, and Wilson, 2006). Students often complain of imbalance in the teacher-made test where attention is paid to minute details in the examination or that emphasis was placed in

certain portions of the content. Either too many items are drawn from an aspect that was given scanty attention during teaching process or an aspect that was not covered in the class receives high weighting when it comes to the test or examination. This is because of the non-use of the table of specification, though table of specifications does not promise a perfectly equitable distribution of weight but it greatly improves the content validity of a teacher-made test (Denga, 2003).

The construction and use of table of specifications serve as blueprint or guide that provides a guide and dictates the number of items that must be administered to measure the subject matter content in each of the topics at each of the cognitive levels. It thus ensures the adequate coverage of both the subject matter content and the different levels of human cognitive behaviour. Therefore, it is one of the most effective empirical means within the teacher's reach of ensuring or building in a high level of content validity for a classroom test.

RECOMMENDATIONS

A classroom test provides teachers with essential information that they can use to make decisions about instructions, students learning and student grades. Based on the issues discussed, the following recommendations are proffered;

1. School administrators could encourage teachers to construct a test blueprint before setting a test so as to improve on the validity of the teacher evaluation.

2. Regional Pedagogic Inspectors could organize seminars /workshops to train teachers on how to construction table of specification.

3. Teachers' training colleges and faculties could emphasize on the importance of the table of specification on test construction.

4. Test developers could be reminded to always use tables of specification when setting their proposed questions to the GCE board.

REFERENCES

[1] Akem, J.A. & Agbe. N. N. (2003). *Measurement and Evaluation in Educational Psychology*. The Return presses.

[2] Akem. J.A. (2006). *Evaluation Techniques in schools and colleges." A Handbook for teachers*. Selfers pub.

[3] Alade, O. M. and Igbinosa, V. O. (2014). Table of Specification and Its Relevance in Educational Development Assessment.

European Journal of Educational and Development Psychology 2(1).

[4] Bloom, B. S. (ed.) (1956). Taxonomy of educational objectives, Handbook 1: The cognitive domain. Mckay.

[5] Denga, D.I. (2003). Educational measurement continuous assessment and psychological testing. (3rd ed.). Rapid Educational Publishers Ltd.

[6] Ehiagwina Osebhohiemen (2019). Use of Table of Specification in Construction of Teacher-Made Achievement Test in Mathematics in the Primary and Secondary Schools. *Melting Pot, 5 (2)*

[7] Fives, Helenrose and DiDonato-Barnes, Nicole (2013) Classroom Test Construction: The Power of a Table of Specifications. *Practical Assessment, Research, and Evaluation: 18 (3).*

[8] Gregory (2006). *Measurement and Assessment in Education.* PEARSON

[9] Grondlund, N. E. (2006). Assessment of students' achievement (8th ed.). PEARSON.

[10] Gronlund, N. E & Linn, R. L. (2000). *Measurement and Assessment in Teaching.* Merrill

[11] Helenrose & DiDonato-Barnes, Nicole (2013). Classroom Test Construction: The Power of a Table of Specifications. *Practical Assessment, Research & Evaluation, 18 (3).*

[12] Joshua, M. T. (2005). *Fundamental test and measurement in education*. The University of Calabar Press

[13] Linn, R. L. & Gronlund, N. E. (2000). *Measurement and assessment in teaching*. Columbus, OH: Merrill.

[14] Moore, K.D. (2001). *Classroom Teaching Skills*. McGraw – Hill Companies Inc. NY.

[15] Nenty, H. J. & Imo, U. E. (2004). *Principles of test construction.* Helimo Associate.

[16] Nenty, H.J. & Umoinyang, I. E. (2007). *Principles of test construction.* Helimo Associate, Capes Publishers

[17] Notar, C.E., Zuelke, D. C., Wilson, J. D. & Yunker, B. D. (2004). The table of specifications: Insuring accountability in teacher made tests. *Journal of Instructional Psychology*, 31, 115- 129

[18] Okpala, N. N, Onacha. C.O & Oyedji, O.A (2003). *Measurement in Education*. Edo State, Stiring Hordon Pub Ltd.

[19] Reynolds, C. R., Livingston, R. B., & Wilson, V. (2006). *Measurement and Assessment in Education*. Pearson.

[20] Wolnring, S. & Wikstron, C. (2010). The concept of validity in theory and practice. Assessment in Education: Principles, *Policy & Practice,* 17(1).

[21] Zuelke, D.C., Wilson. J.D. & Yunker, B.D (2004). The table of Specification: Insuring accountability in teacher made tests. *Journal of Instructional Psychology*, 31, 115-129